# METHOD AND APPARATUS FOR TWO-PORT ALLPASS COMPENSATION OF POLARIZATION MODE DISPERSION

### Field of the Invention

5              The present invention relates generally to optical fiber transmission systems and, more particularly, to methods and apparatus for measuring and controlling polarization mode dispersion (PMD) in such optical fiber transmission systems.

### Background of the Invention

10              Optical communication systems increasingly employ wavelength division multiplexing (WDM) techniques to transmit multiple information signals on the same fiber at different optical frequencies. WDM techniques are being used to meet the increasing demands for improved speed and bandwidth in optical transmission applications, including fiber optic communication systems.

15              Polarization mode dispersion (PMD) is a fundamental problem in single-mode fiber optic communication systems that has limited the channel capacity. Fiber asymmetry gives rise to random birefringence effects, resulting in differential group delay (DGD) between the two principle polarization modes. Wavelength dependent coupling of the modes then leads to polarization mode dispersion, which becomes the dominant

20  limiting factor of transmission rate as bandwidth increases.

              A need therefore exists for improved techniques for compensating for polarization mode dispersion in order to further improve the efficiency and channel capacity of high-speed fiber optic communication.

### 25  Summary of the Invention

              Generally, a method and apparatus are disclosed for compensating for polarization mode dispersion using cascades of all-pass filters and directional couplers. The disclosed PMD compensator adjusts the coefficients of an adaptive filter structure involving all-pass filters and directional couplers based on a minimized cost function. In

30  one implementation, a stochastic gradient algorithm, also referred to as the least mean square (LMS) algorithm, is employed to sequentially reduce the value of the cost function by the method of steepest descent. In another implementation, convergence is improved by employing a Newton algorithm that uses second derivatives to accelerate convergence.

A more complete understanding of the present invention, as well as further features and advantages of the present invention, will be obtained by reference to the following detailed description and drawings.

<br>

5     **Brief Description of the Drawings**

FIG. 1 is a schematic block diagram of a conventional optical receiver that employs a PMD compensator;

FIG. 2 is a schematic block diagram illustrating of a conventional optical fiber transmission system that employs a PMD compensator;

10     FIG. 3 is a schematic block diagram of the PMD measuring apparatus of FIG. 1;

FIG. 4 is a schematic block diagram of a polarization mode dispersion compensator for use in an optical receiver;

FIG. 5 illustrates an exemplary model of a PMD generator; and

15     FIG. 6 illustrates an exemplary model of a PMD compensator.

<br>

**Detailed Description**

The present invention provides a method and apparatus for compensating for polarization mode dispersion using cascades of all-pass filters and directional couplers.

20     The disclosed PMD compensator adjusts the coefficients of an adaptive filter structure involving all-pass filters and directional couplers based on a minimized cost function. Initially, the phase and amplitude of polarization components are evaluated in order to characterize the PMD of channels and adjust tunable PMD compensators. FIG. 1 is a schematic block diagram of a conventional optical receiver 100 that employs a PMD

25     compensator 400, discussed below in conjunction with FIG. 4. Generally, the PMD compensator 400 employs a PMD measuring apparatus 300, discussed further below in conjunction with FIG. 3, to measure the PMD and an adaptive optical filter 110 to compensate for the measured PMD. The PMD compensator 400 may be integrated with or in the vicinity of the optical receiver 100.

30     As shown in FIG. 1, an information signal is received by the optical receiver 100 over an optical fiber 105. The PMD compensator 400 measures the phase and magnitude of the polarization components and generates a feedback signal for

controlling one or more adaptive optical filters 300 to reduce polarization mode dispersion.

FIG. 2 is a schematic block diagram illustrating of a conventional optical fiber transmission system 200 that employs the PMD compensator 400. A test signal 222

5    is added to a data signal 221 in the form of a comb of tones with known relative magnitude and phase relationships. The tones can be equally spaced. The spacing $\delta$ f between the tones may be on the order of 2.5 to 5 GHz. The tones provide test signal information with a higher signal-to-noise ratio at the receiver than a pseudo-random bit sequence. There are many ways to add such a test signal 222, including modifying the

10   data format such that tones with the desired frequency spacing appear in the output spectrum 223. The combined data signal and test signal is fed to a laser/modulator combination 224 to generate a wavelength channel (as a channel in a wavelength division multiplexed (WDM) transmission system). The channel is combined with other channels in an optical multiplexer 225 to form a composite light signal. The composite light signal

15   is then transmitted via optical transmission path 213.

At a receiver, demultiplexer 226, preferably at the downstream end of optical transmission path 213, separates the wavelength channels. Optical coupler 227 provides a wavelength channel to the polarization mode dispersion compensator (PMDC) 110 (the adaptive optical filter). Coupler 227 also provides light to the PMD measuring

20   apparatus 300. The channel estimate made at PMD measuring apparatus 300 generates an adaptive PMD correction signal to control PMDC 110. Receiver (RX) 100 receives the compensated wavelength channel signal. The RF detectors (not shown in FIG. 2) of the channel estimate signal analyzer 300 need only have enough bandwidth to accommodate $\delta$ f and not the whole signal bandwidth. The problem is to determine the relative phase

25   between each pair of tones when they are all present at the detector.

FIG. 3 schematically illustrates an exemplary PMD measuring apparatus 300 for analyzing the test signal downstream. Each polarization is split into a separate path by a polarization beam splitter (PBS) and polarization controller (PC) 301. The PC 301 allows the power in the x- and y-outputs to be controlled so that all of the power is not

30   in one output or the other. One polarization of the output of the PBS/PC is flipped by a 90 degree rotation 302 so that the outputs going into the 3 dB couplers 303 have the same polarization, i.e., either TE or TM. After the 3 dB coupler, one portion of each

polarization is analyzed by a tunable narrowband filter (NBF) 304 to obtain the magnitude across the channel (detectors X 308, Y 309), and $\phi$ the relative phase between polarization (detector 310) derived via 3 dB coupler 303, while the other portion is transmitted through a tunable all-pass filter (APF) 305 before being detected.

5          Both polarizations see the same narrowband filter but in counterpropagating directions. The filters are tuned, for example, by thermo-optic phase shifters. The APF's are identical in principle, but any variations can be handled by calibration. Each APF is designed to provide a very sharp transition in its phase response from 0 to $2\pi$ near resonance. On-resonance, the phase is $\pi$. Off-resonance, the phase is

10         ideally 0 or $2\pi$. As the resonant frequency is shifted via a phase shifter in the feedback path, the phase response is translated across the channel spectrum and the RF detectors X 306 and Y 307 record different linear combinations of beats between adjacent tones. An RF reference 313 is obtained from a light signal tapped before the polarization beam splitter 301 and fed to an RF detector and phase locked loop incorporating a voltage

15         controlled oscillator 314 to develop an RF reference signal 313. Blocks 311 and 312 measure the phase of the X and Y components as detected by 306 and 307 with respect to the RF reference 313. For a detailed discussion of the PMD measuring apparatus 300, see United States Patent Application Serial Number 10/180,842, entitled "Apparatus and Method for Measurement and Adaptive Control of Polarization Mode Dispersion in

20         Optical Fiber Transmission Systems," incorporated by reference herein.

According to another aspect of the invention, adaptive algorithms are provided for two-channel (two-input/two-output) structures consisting of multiple cascades of all-pass filters and directional couplers. While this exemplary architecture is employed for compensation of polarization mode dispersion, the results apply more

25         generally to this class of filters, as would be apparent to a person of ordinary skill in the art. For a detailed discussion of an application of this architecture to PMD compensation using multiple cascades of all-pass filters and directional couplers, see C.K. Madsen, "Optical All-Pass Filters for Polarization Mode Dispersion Compensation," Optics Letters, Vol. 25, No. 12, 878-80 (June, 2000), incorporated by reference herein.

30         FIG. 4 is a schematic block diagram of a conventional polarization mode dispersion compensator 400 for use in an optical receiver, such as the optical receiver 100 of FIG. 1. As shown in FIG. 4, the PMD compensator 400 is a two-channel (two-

input/two-output) structure consisting of multiple cascades of all-pass filters and directional couplers.  In particular, the PMD compensator 400 includes a polarization beam splitter 410, two pair of multistage all-pass filters 420-1, 420-2, 440-1, 440-2 and two directional couplers 430, 450, that all operate in a known manner.

5          Models for PMD Generation and Compensation

FIG. 5 illustrates an exemplary model of a PMD generator, where the two dimensional polarization z-transform vector $x(z)$ generated by the transmitter produces the fiber output polarization vector $y(z)$ at the receiver.  $A_{opt}$ 540 and $B_{opt}$ 520 are multistage all-pass filters and

10
$$T = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 & -j \\ -j & 1 \end{pmatrix} \qquad (1)$$

is a unitary matrix representing a directional coupler 510, 530 with parameter value 0.5 and $j = \sqrt{-1}$.  In this manner, the simplified PMD compensator 600 of FIG. 6 has the ideal capability of complete compensation, i.e., the compensator output $\hat{x}(z) = x(z)$, when $A = A_{opt}$ and $B = B_{opt}$.  This will allow observations about the behavior of parameter

15      estimation algorithms to be made without being overly burdened by the complications of real PMD.

The multistage all-pass filter $A$ 540 is mathematically expressed as:

$$A(z) = \prod_{p=1}^{P} A_p(z), \qquad (2)$$

where $P$ is the number of stages and the response of each stage is written as follows:

20
$$A_p(z) = \frac{z^{-1} - a_p^*}{1 - a_p z^{-1}}, \qquad (3)$$

where $a_p$ is a complex number that specifies the pole location and * denotes complex conjugate.  (The complementary zero location is at $1/a_p^*$.)  The sections $A_p(z)$ are all-pass functions so that $|A_p(z)| = 1$ at all frequencies.

Likewise, a Q-stage filter $B$ 520 is defined as

25
$$B(z) = \prod_{q=1}^{Q} B_q(z), \qquad (4)$$

where $Q$ is the number of stages and the response of each stage is written as follows:

$$B_q(z) = \frac{z^{-1} - b_q^*}{1 - b_q z^{-1}} \tag{5}$$

Cascading the above in the order of FIG. 5 and multiplying out the 2x2 matrices gives the overall response of the PMD model:

5

$$y = \frac{1}{2} \begin{bmatrix} A_{opt}^*(B_{opt}^* - 1) & jA_{opt}^*(B_{opt}^* + 1) \\ j(B_{opt} + 1) & -(B_{opt}^* - 1) \end{bmatrix} x \tag{6}$$

This shows how the input polarization signals are coupled to the output signals, as determined by the all-pass model filters $A_{opt}$ 540 and $B_{opt}$ 520. When $A_{opt} = B_{opt} = -1, y = x$ and there is no distortion.

Now consider the response of the PMD compensator 600 of FIG. 6, which

10   is similarly calculated as:

$$\hat{x} = \frac{1}{2} \begin{bmatrix} A(B-1) & -j(B+1) \\ -jA(B+1) & -(B-1) \end{bmatrix} y. \tag{7}$$

Again, note that $\hat{x} = y$ for $A = B = -1$. Also note that for $A = A_{opt}$ and $B = B_{opt}$, the matrices in Equations (6) and (7) are conjugate transpose pairs; in fact, it can be easily demonstrated that in this case they are unitary so that their product is the identity

15   matrix and $\hat{x} = x$. Thus, the compensator completely removes the modeled PMD for $A = A_{opt}$ and $B = B_{opt}$.

The difference between the ideal input signal vector $x$ and its recovered estimate $\hat{x}$ forms an error signal vector

$$e(z) = x(z) - \hat{x}(z) \tag{8}$$

20   which should be small. In other words, the error signal vector, e(z), (having two values, one for each principle polarization mode) characterizes the difference between the signal that was transmitted, x(z), (which is known) and the response, $\hat{x}(z)$, of the PMD compensator. Here, it is assumed that compensation is performed over a discrete set of $K$ representative frequencies $z_k, k = 1, \ldots, K$, and form the weighted cost function:

25

$$J = \sum_{k=1}^{K} \left[ \lambda_1 |e_1(z_k)|^2 + \lambda_2 |e_2(z_k)|^2 \right]. \tag{9}$$

Equation (9) looks at the two components of the error signal vector, e(z), as functions of frequency ($z_k$). Generally, Equation (9) determines a weighted sum of the squared error, that is summed over K frequencies.

As discussed hereinafter, the present invention provides adaptive
5   algorithms for adjusting $A$ and $B$ to minimize this cost function, J. The complex (amplitude and phase) measurements of the optical PMD compensator output $\hat{x}$ for each frequency $z_k$ are provided by the detector 310 of FIG. 3. Generally, such measurements can be obtained using a tunable narrowband optical filter and various other optical components to render information from energy detector measurements. For a more
10  detailed discussion of the detector 310 for measuring the complex (amplitude and phase) measurements of the optical PMD compensator output $\hat{x}$ for each frequency $z_k$, see United States Patent Application Serial Number 10/180,842, entitled "Apparatus and Method for Measurement and Adaptive Control of Polarization Mode Dispersion in Optical Fiber Transmission Systems," incorporated by reference herein.

15                                  Adaptive Algorithms

According to another aspect of the invention, two exemplary algorithms are presented for adapting the coefficients of a two-channel adaptive filter structure involving two all-pass filters and two directional couplers. Both exemplary adaptive algorithms minimize the cost function of equation (9). The first adaptive algorithm is a stochastic
20  gradient algorithm, also referred to as the least mean square algorithm, and sequentially reduces the value of $J$ by the method of steepest descent. For the simplified model of FIGS. 5 and 6, the effects of random signals or noise are not considered, so the LMS algorithm in this case operates on deterministic signals (thus, the LMS algorithm is not really forming a stochastic gradient).

25                     As is well known, the LMS algorithm has the disadvantage of slow convergence for some applications in which the effect of the adjustable coefficients is closely coupled. As shown hereinafter, such behavior results for the all-pass PMD compensation context of the present invention. Therefore, a Newton algorithm is also derived for this application that makes use of second derivatives to accelerate
30  convergence. For both algorithms, it will be necessary to calculate derivatives of the error signals $e_1$ and $e_2$ in (9).

## LMS Algorithm

The LMS algorithm adapts the (complex) coefficients of the PMD compensator all-pass filters employed by equations (3) and (5) as follows:

$$a_p(n+1) = a_p(n) - \mu \frac{\partial J}{\partial a_p} \tag{10a}$$

$$b_q(n+1) = b_q(n) - \mu \frac{\partial J}{\partial b_q} \tag{10b}$$

where $\mu$ is the step size and $\dfrac{\partial J}{\partial c}$ denotes the complex derivative of $J$ with respect to the complex variable $c$, defined as:

$$\frac{\partial J}{\partial c} = \frac{\partial}{\partial \Re(c)} + j \frac{\partial}{\partial \Im(c)}, \tag{11}$$

where $\Re(c)$ and $\Im(c)$ are, respectively, the real and imaginary parts of $c = \Re(c) + j\Im(c)$.

As previously indicated, these updates are deterministic for the simplified model. However, in the case of random signals, this also provides instantaneous gradient updates which are subsequently smoothed by the recursive nature of the algorithm, as controlled by the selection of $\mu$. In general, the selection of step size value must take into account the competing objectives of rapid convergence/tracking, and maintaining stability (and smoothing in the case of random signals).

Differentiating equation (9) with respect to $a_p$, it can be shown that:

$$\frac{\partial J}{\partial a_p^r} = 2\sum_{k=1}^{K} \Im\{G_p(z_k)\}F_1(z_k) \tag{12a}$$

$$\frac{\partial J}{\partial a_p^i} = 2\sum_{k=1}^{K} \Re\{G_p(z_k)\}F_1(z_k), \tag{12b}$$

where $a_p^r$ and $a_p^i$ are, respectively, the real and imaginary parts of $a_p \equiv a_p^r + j a_p^i$,

$$G_p(z) \equiv \left(\frac{1}{z - a_p}\right), \tag{13}$$

and

$$F_1(z) \equiv \lambda_1 \Im\{A(z)[B(z) - 1] y_1(z) e_1^*(z)\} \\ - \lambda_2 \Re\{A(z)[B(z) + 1] y_1(z) e_2^*(z)\} \tag{14}$$

is a real valued function, independent of the index $p$. Combining equations (12a) and (12b), the complex derivative can be more compactly written as

$$\frac{\partial J}{\partial a_p} = j2\sum_{k=1}^{K} G_p^*(z_k) F_1(z_k).$$ (15)

The complex derivative can be alternatively calculated by considering $a_p$ and $a_p^*$ to be independent variables and differentiating with respect to $a_p^*$. However, the derivatives are separately derived with respect to the real and imaginary parts of $a_p$ because they will be needed in the next section to calculate the second derivatives.

Likewise, for $B$, the following is obtained:

$$\frac{\partial J}{\partial b_q^r} = 2\sum_{k=1}^{K} \Im\{H_q(z_k)\}F_2(z_k)$$ (16a)

$$\frac{\partial J}{\partial b_q^i} = 2\sum_{k=1}^{K} \Re\{H_q(z_k)\}F_2(z_k)$$ (16b)

where

$$H_q(z) \equiv \left(\frac{1}{z - b_q}\right)$$ (17)

and

$$F_2(z) \equiv \lambda_1 \Im\{B(z)[A(z)y_1(z) - jy_2(z)]e_1^*(z)\} \\ - \lambda_2 \Re\{B(z)[A(z)y_1(z) - jy_2(z)e_2^*(z)]\},$$ (18)

and

$$\frac{\partial J}{\partial b_q} = j2\sum_{k=1}^{K} H_q^*(z_k) F_2(z_k).$$ (19)

A composite coefficient vector is defined as:

$$w = \begin{bmatrix} a \\ b \end{bmatrix}$$ (20)

where $a \equiv [a_1\ a_2\ \ldots\ a_p]^T$ and $b \equiv [b_1\ b_2\ \ldots\ b_Q]^T$, then equation (10) can be compactly written as:

$$w(n+1) = w(n) - \mu\nabla(J),$$ (21)

where

$$\nabla(J) \equiv \left[ \frac{\partial J}{\partial \mathbf{a}^T} \frac{\partial J}{\partial \mathbf{b}^T} \right]^T. \qquad (22)$$

is the $(P+Q) \times 1$ complex gradient of $J$ with respect to w, and

$$\frac{\partial J}{\partial \mathbf{a}^T} \equiv \left[ \frac{\partial J}{\partial a_1} \quad \frac{\partial J}{\partial a_2} \quad \cdots \quad \frac{\partial J}{\partial a_P} \right] \qquad (23a)$$

$$\frac{\partial J}{\partial \mathbf{b}^T} \equiv \left[ \frac{\partial J}{\partial b_1} \quad \frac{\partial J}{\partial b_2} \quad \cdots \quad \frac{\partial J}{\partial b_Q} \right] \qquad (23b)$$

5         It has been observed that the LMS algorithm demonstrates slow convergence, which can be solved using the Newton algorithm to accelerate convergence, discussed hereinafter.

### Newton Algorithm

Newton's method involves multiplying the gradient by the inverse matrix of
10    second derivatives, or Hessian. It has been found that it is not generally possible to apply this to the complex LMS algorithm defined by equation (21), whereby the Hessian is a complex matrix, even in the simplest case with a single complex coefficient. Therefore, the real and imaginary components of the weight update, as in equations (12) and (16), must be dealt with separately.

15         Taking second derivatives of equation (9) with respect to the real and imaginary parts of $a_p$ and $b_p$, it can be shown that:

$$\frac{\partial^2 J}{\partial a_q^r \partial a_p^r} = 2 \sum_{k=1}^K \left\{ \Im[G_q(z_k)]\Im[G_p(z_k)]G_{11}(z_k) + \Im[G_p^2(z_k)]F_1(z_k)\delta_{pq} \right\} \quad (24a)$$

$$\frac{\partial^2 J}{\partial a_q^i \partial a_p^r} = 2 \sum_{k=1}^K \left\{ \Re[G_q(z_k)]\Im[G_p'(z_k)]G_{11}(z_k) + \Re[G_p^2(z_k)]F_1(z_k)\delta_{pq} \right\} \quad (24b)$$

$$\frac{\partial^2 J}{\partial a_q^i \partial a_p^i} = 2 \sum_{k=1}^K \left\{ \Re[G_q(z_k)]\Re[G_p(z_k)]G_{11}(z_k) - \Im[G_p^2(z_k)]F_1(z_k)\delta_{pq} \right\} \quad (24c)$$

20

$$\frac{\partial^2 J}{\partial b_q^r \partial b_p^r} = 2 \sum_{k=1}^K \left\{ \Im[H_q(z_k)]\Im[H_p(z_k)]G_{22}(z_k) + \Im[H_p^2(z_k)]F_1(z_k)\delta_{pq} \right\} \quad (25a)$$

$$\frac{\partial^2 J}{\partial b_q^i \partial b_p^r} = 2 \sum_{k=1}^K \left\{ \Re[H_q(z_k)]\Im[H_p(z_k)]G_{22}(z_k) + \Re[H_p^2(z_k)]F_1(z_k)\delta_{pq} \right\} \quad (25b)$$

$$\frac{\partial^2 J}{\partial b_q^i \partial b_p^i} = 2\sum_{k=1}^{K} \left\{ \Re[H_q(z_k)]\Re[H_p(z_k)]G_{22}(z_k) - \Im[H_p^2(z_k)]F_1(z_k)\delta_{pq} \right\} \quad (25c)$$

$$\frac{\partial^2 J}{\partial b_q^r \partial a_p^r} = \frac{\partial^2 J}{\partial a_p^r \partial b_q^r} = 2\sum_{k=1}^{K} \Im[H_q(z_k)]\Im[G_p(z_k)]G_{12}(z_k) \quad (26a)$$

$$\frac{\partial^2 J}{\partial b_q^i \partial a_p^r} = \frac{\partial^2 J}{\partial a_p^r \partial b_q^i} = 2\sum_{k=1}^{K} \Re[H_q(z_k)]\Im[G_p(z_k)]G_{12}(z_k) \quad (26b)$$

$$\frac{\partial^2 J}{\partial b_q^r \partial a_p^i} = \frac{\partial^2 J}{\partial a_p^i \partial b_q^r} = 2\sum_{k=1}^{K} \Im[H_q(z_k)]\Re[G_p(z_k)]G_{12}(z_k) \quad (26c)$$

$$\frac{\partial^2 J}{\partial b_q^i \partial a_p^i} = \frac{\partial^2 J}{\partial a_p^i \partial b_q^i} = 2\sum_{k=1}^{K} \Re[H_q(z_k)]\Re[G_p(z_k)]G_{12}(z_k), \quad (26d)$$

where $G_p$ and $H_q$ are defined, respectively, in (13) and (17), $\delta_{pq} = 1$ for $p-q$ and 0 otherwise, and

$$G_{11}(z) \equiv \lambda_1 \Re\{2A(z)[B(z)-1]y_1(z)e_1^*(z) + \left|[B(z)-1]y_1(z)\right|^2\}$$

$$+ \lambda_1 \Im\{2A(z)[B(z)+1]y_1(z)e_2^*(z) + j\left|[B(z)+1]y_1(z)\right|^2\} \quad (27a)$$

$$G_{22}(z) \equiv \lambda_1 \Re\{2B(z)[A(z)y_1(z) - jy_2(z)]e_1^*(z) + \left|A(z)y_1(z) - jy_2(z)\right|^2\}$$

$$+ \lambda_2 \Im\{2B(z)[A(z)y_1(z) - jy_2(z)]e_2^*(z) + j\left|A(z)y_1(z) - jy_2(z)\right|^2\} \quad (27b)$$

$$G_{12}(z) \equiv \lambda_1 \Re\{2A(z)B(z)y_1(z)e_1^*(z) + A*(z)B(z)[B*(z)-1][A(z)y_1(z) - jy_2(z)]y_1^*(z)\}$$

$$+ \lambda_1 \Im\{2A(z)B(z)y_1(z)e_2^*(z) + jA*(z)B(z)[B*(z)+1][A(z)y_1(z) - jy_2(z)]y_1^*(z)\} \quad (27c)$$

are real functions.

It is noted that as $e_1 \to 0$ and $e_2 \to 0$, $G_{11}$ and $G_{22}$ are non-negative, and $F_1 \to 0$ and $F_2 \to 0$. Therefore, as the error goes to zero, equations (24) and (25) show that the diagonal terms

$$\frac{\partial^2 J}{\partial a_p^r \partial a_p^r} \quad \frac{\partial^2 J}{\partial a_p^i \partial a_p^i} \quad \frac{\partial^2 J}{\partial b_q^r \partial b_q^r} \quad \frac{\partial^2 J}{\partial b_q^i \partial b_q^i} \quad (28)$$

are all non-negative, thus establishing a quadratic minimum.

The $P \times P$ Hessian sub-matrix is now assembled from equation (24a) as

$$\frac{\partial^2 J}{\partial \mathbf{a}^r \partial \mathbf{a}^{rT}} = \begin{bmatrix} \dfrac{\partial^2 J}{\partial a_1^r \partial a_1^{rT}} & \dfrac{\partial^2 J}{\partial a_1^r \partial a_2^{rT}} & \cdots & \dfrac{\partial^2 J}{\partial a_1^r \partial a_P^{rT}} \\ \dfrac{\partial^2 J}{\partial a_2^r \partial a_1^{rT}} & \dfrac{\partial^2 J}{\partial a_2^r \partial a_2^{rT}} & \cdots & \dfrac{\partial^2 J}{\partial a_2^r \partial a_P^{rT}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 J}{\partial a_P^r \partial a_1^{rT}} & \dfrac{\partial^2 J}{\partial a_P^r \partial a_2^{rT}} & \cdots & \dfrac{\partial^2 J}{\partial a_P^r \partial a_P^{rT}} \end{bmatrix} \tag{29}$$

Likewise, $\dfrac{\partial^2 J}{\partial \mathbf{a}^i \partial \mathbf{a}^{iT}}$ and $\dfrac{\partial^2 J}{\partial \mathbf{a}^r \partial \mathbf{a}^{iT}} = \dfrac{\partial^2 J}{\partial \mathbf{a}^i \partial \mathbf{a}^{rT}}$ are formed. Then, defining the $2P \times 1$ coefficient vector as:

$$\mathbf{a} \equiv \begin{bmatrix} \mathbf{a}^r \\ \mathbf{a}^i \end{bmatrix} \tag{30}$$

5    where $\mathbf{a}^r \equiv [a_1^r \quad a_2^r \quad \ldots \quad a_P^r]$ and $\mathbf{a}^i \equiv [a_1^i \quad a_2^i \quad \ldots \quad a_P^i]$, we form the $2P \times 2P$ Hessian matrix:

$$\frac{\partial^2 J}{\partial \mathbf{a} \partial \mathbf{a}^T} = \begin{bmatrix} \dfrac{\partial^2 J}{\partial \mathbf{a}^r \partial \mathbf{a}^{rT}} & \dfrac{\partial^2 J}{\partial \mathbf{a}^r \partial \mathbf{a}^{iT}} \\ \dfrac{\partial^2 J}{\partial \mathbf{a}^i \partial \mathbf{a}^{rT}} & \dfrac{\partial^2 J}{\partial \mathbf{a}^i \partial \mathbf{a}^{iT}} \end{bmatrix}. \tag{31}$$

Similarly, from equations (25) and (26), the $2Q \times 2Q$ matrix is obtained as follows:

10

$$\frac{\partial^2 J}{\partial \mathbf{b} \partial \mathbf{b}^T} = \begin{bmatrix} \dfrac{\partial^2 J}{\partial \mathbf{b}^r \partial \mathbf{b}^{rT}} & \dfrac{\partial^2 J}{\partial \mathbf{b}^r \partial \mathbf{b}^{iT}} \\ \dfrac{\partial^2 J}{\partial \mathbf{b}^i \partial \mathbf{b}^{rT}} & \dfrac{\partial^2 J}{\partial \mathbf{b}^i \partial \mathbf{b}^{iT}} \end{bmatrix} \tag{32}$$

and the $2Q \times 2P$ matrix is expressed as follows:

$$\frac{\partial^2 J}{\partial \mathbf{b} \partial \mathbf{a}^T} = \left[ \frac{\partial^2 J}{\partial \mathbf{a} \partial \mathbf{b}^T} \right]^T = \begin{bmatrix} \dfrac{\partial^2 J}{\partial \mathbf{b}^r \partial \mathbf{a}^{rT}} & \dfrac{\partial^2 J}{\partial \mathbf{b}^r \partial \mathbf{a}^{iT}} \\ \dfrac{\partial^2 J}{\partial \mathbf{b}^i \partial \mathbf{a}^{rT}} & \dfrac{\partial^2 J}{\partial \mathbf{b}^i \partial \mathbf{a}^{iT}} \end{bmatrix} \tag{33}$$

where

$$\mathbf{b} \equiv \begin{bmatrix} \mathbf{b}^r \\ \mathbf{b}^i \end{bmatrix}. \tag{34}$$

The above results are consolidated and the Newton algorithm is then formulated. First, a composite real coefficient vector is defined as follows:

$$w \equiv [a^{rT} \quad a^{iT} \quad b^{rT} \quad b^{iT}]^T = \begin{bmatrix} a \\ b \end{bmatrix} \tag{35}$$

and the composite $(2P+2Q) \times (2P+2Q)$ Hessian matrix is formed as follows:

$$5 \qquad H = \frac{\partial^2 J}{\partial w \partial w^T} = \begin{bmatrix} \dfrac{\partial^2 J}{\partial a \partial a^T} & \dfrac{\partial^2 J}{\partial a \partial b^T} \\ \dfrac{\partial^2 J}{\partial b \partial a^T} & \dfrac{\partial^2 J}{\partial b \partial b^T} \end{bmatrix}. \tag{36}$$

Similarly, using equations (12) and (16), the real gradient of $J$ can be assembled with respect to the composite real coefficient vector (35) as follows:

$$\nabla(J) \equiv \begin{bmatrix} \dfrac{\partial J}{\partial a^{rT}} & \dfrac{\partial J}{\partial a^{iT}} & \dfrac{\partial J}{\partial b^{rT}} & \dfrac{\partial J}{\partial b^{iT}} \end{bmatrix}^T = \begin{bmatrix} \dfrac{\partial J}{\partial a^T} & \dfrac{\partial J}{\partial b^T} \end{bmatrix}^T. \tag{37}$$

Finally, with the above definitions, the Newton algorithm is written as

10 $\qquad w(n+1) = w(n) - \mu H^{-1} \nabla(J).$ $\qquad\qquad$ (38)

Generally, it has been observed that with the Newton algorithm, the error converges more quickly and steadily to zero, relative to the LMS algorithm which takes a longer time to converge completely. Even for 10% initialization with the Newton algorithm, convergence to zero was rapidly attained, although there may be some potential

15 instability at some intermediate point.

In one variation, the capability of the Newton algorithm is extended by first running the LMS algorithm for a small number of samples and then using the coefficient values to initialize the Newton algorithm. It is noted that both the LMS and Newton algorithms are susceptible to getting stuck in a local minimum. The theoretical basis for

20 this phenomenon is related to the nature of the MSE gradient, which for the all-pass structure turns out to have components that only differ by a function that depends on the coefficient value. Therefore, if any two coefficients ever reach the same value they will remain locked together, thereby leading to the local minimum. One way around this problem is to compute multiple solutions starting with a suitable number of initial guesses

25 spaced out over the feasible parameter space.

As is known in the art, the methods and apparatus discussed herein may be distributed as an article of manufacture that itself comprises a computer readable medium

having computer readable code means embodied thereon.  The computer readable program code means is operable, in conjunction with a computer system, to carry out all or some of the steps to perform the methods or create the apparatuses discussed herein.  The computer readable medium may be a recordable medium (e.g., floppy disks, hard drives, compact disks such as DVD, or memory cards) or may be a transmission medium (e.g., a network comprising fiber-optics, the world-wide web, cables, or a wireless channel using time-division multiple access, code-division multiple access, or other radio-frequency channel). Any medium known or developed that can store information suitable for use with a computer system may be used. The computer readable code means is any mechanism for allowing a computer to read instructions and data, such as magnetic variations on a magnetic media or height variations on the surface of a compact disk, such as a DVD.

It is to be understood that the embodiments and variations shown and described herein are merely illustrative of the principles of this invention and that various modifications may be implemented by those skilled in the art without departing from the scope and spirit of the invention.